ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Physics Procedia 22 (2011) 471 – 477

Physics Procedia

2011 International Conference on Physics Science and Technology (ICPST 2011)

# Content-Based Image Retrieval Research

Guoyong Duan[a], Jing Yang[a], Yilong Yang[a]

[a] *Department of Computer Science and Information, Guizhou University, Guiyang, 550025, China*

**Abstract**

With the development of multimedia technology, the rapid increasing usage of large image database becomes possible. To carry out its management and retrieval, Content-Based Image Retrieval (CBIR) is an effective method. This paper shows the advantage of content-based image retrieval system, as well as key technologies. Compare to the shortcoming that only certain one feature is used in the traditional system, this paper introduces a method that combines color, texture and shape for image retrieval and shows its advantage. Then this paper focuses on the feature extraction and representation, several commonly used algorithms and image matching methods.



*PACS:* Type pacs here, separated by semicolons ;
*Keywords:* CBIR, Image Segmentation, Feature extraction, Color, Texture, Shape

* Corresponding author. Tel.:+86-13885167705
E-mail address: yangjing6646@yahoo.com.hk

* Supported by Guizhou fund projects Qianke-heJzi[2010]2094, Guiyang city bureau of industrial research projects[2010]zhukegonghetongzi 1-75, Supported by Guiyang city bureau of Science and technology plan projects [2011]zhukehetong2011201da–4-2

 doi:10.1016/j.phpro.2011.11.073

## 1. Introduction

Content-based image retrieval (CBIR), which belongs to a research field of image analysis, also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR). [1] The key technologies image retrieval include: image feature extraction, feature-based similarity calculation, semantically relevance feedback and image acquisition.[2,3] It relates to machine vision, pattern recognition, database technology and information retrieval studies. For general-purpose image retrieval, the main features used to retrieve are: color, texture, shape, figure, layout, etc. among which color, texture, shape are particularly prevalent.[4,5]

## 2. Image segmentation

With the introduction of GrabCut algorithm, the image processing can eliminate the influence of background. First, the GrabCut generates a basic "hard partition" through interactive segmentation algorithm. Then it acquires a ideal segmentation by using Border Matting around the hard partition boundaries. "Hard partition" improves GrabCut, where improvements are made in three areas: first, it uses Gaussian mixture models (GMM) to replace the histogram, extending gray-level image to color image; second, it uses parameter estimates and evolutional iteration algorithm to replace minimum estimated to achieve the energy minimization; third, it reduces the job requirements by using non-fully number.

Each GMM (foreground or background) can be regarded as a K-dimensional covariance (usually K = 5). In order to facilitate the processing of GMM optimization, the vector $\kappa=(\kappa_1,\ldots,\kappa_n,\ldots,\kappa_N)$. as independent of each pixel GMM (foreground or background) parameters is added, and $\kappa_N \in \{1,2, ..., K\}$, the corresponding pixel on opacity $\alpha_N = 0$ or 1. Gibbs energy function is rewritten as

$$E(\alpha,\kappa,\theta,z)=U(\alpha,\kappa,\theta,z)+V(\alpha,z) \tag{1}$$

Where: $\alpha$ stands for opacity, $\alpha \in [0,1]$, 0 as the background, 1 for the foreground object; $\theta$ stands for histogram of the image of foreground and background, $\theta = \{h(\alpha,z), \alpha = 0,1\}$; $z$ stands for the gray degree array of image values, $z=(z_1,\ldots,z_n,\ \ldots,z_N)$. (1) is mainly affected by the GMM variables $\kappa$. The data items of the GMM color data model can be defined as

$$\mathrm{U}(\alpha,\kappa,\theta,z)=\sum_n(\alpha_n,\kappa_n,\theta,z_n)$$

The parameters of this model are identified as:

$$\theta=\left\{\pi(\alpha_n,\kappa_n),\mu(\alpha_n,\kappa_n),\sum(\alpha,\kappa),\kappa=1,.2,...,K\right\}$$

The smooth entry for color images:

$$\nu(\alpha,z)=\gamma\sum_{\{m,n\}\in C}\left\{\alpha_m\neq\alpha_n\right\}\exp\left\{-\beta\parallel z_m-z_n\parallel^2\right\}$$

The result shows as Fig.1

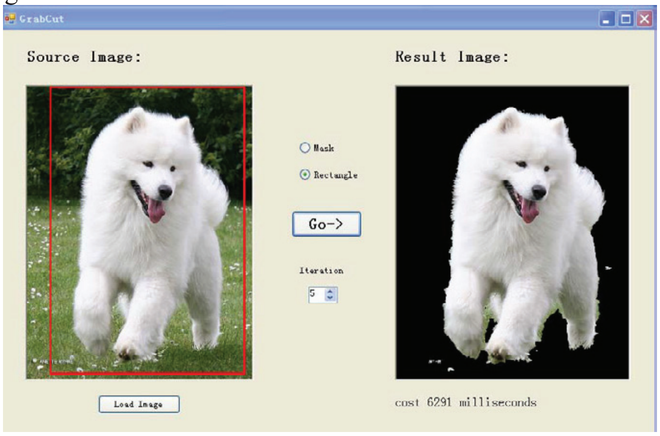


Fig.1 GrabCut algorithm effect

## 3. Color feature extraction

In the current model's colors, RGB color model is a most practical application. RGB color space is convenient for CRT device to display images, and it facilitates image exchange. Digital image are general displayed by RGB color model which is divided into three channels: Red, Green, Blue. The color model reflects the color in each channel on the brightness value respectively. RGB color values can be acquired from the images in general. Conversion from RGB space to HSV space is as follows:

Given the value of RGB color space $(r, g, b)$, where $r, g, b \in [0 \sim 255]$, then we can assume

$$v = \frac{v'}{255}, s = \frac{v' - \min(r, g, b)}{v'}, h = \begin{cases} 5 + b'(r = v') \\ 1 + r'(g = v') \\ 3 + g'(b = v') \end{cases}$$

$$So\ v = \frac{v'}{255}, s = \frac{v' - \min(r, g, b)}{v'}, h = \begin{cases} 5 + b'(r = v') \\ 1 + r'(g = v') \\ 3 + g'(b = v') \end{cases}$$

We use the HSV space color histogram to describe the color feature of the overall image. According to human visual ability, hue(H) is divided into 16 pieces, saturation(S) is divided into 4 pieces and value(V) is divided into 4 pieces. Then non-interval quantization based on different colors combined with the frequency bandwidth of each color, we obtain one-dimensional vector L:

$$L = 16H + 4S + V$$

L is in the range [0, 1,…, 255], and 256 bin one-dimensional histogram is acquired by calculating L.

Let Q as an example of retrieval images, P and R are in the image database images. First, their feature vectors are normalized, namely:

$$L_1 = l_1 / (l_1 + l_2 + ... + l_{256}), L_2 = l_2 / (l_1 + l_2 + ... + l_{256}), ..., L_{256} = l_{256} / (l_1 + l_2 + ... + l_{256})$$

The images P and Q can be represented as following vectors:

$$L(P)=(l_{P1},l_{P2},...,l_{P256}),L(Q)=(l_{Q1},l_{Q2},...,l_{Q256})$$

Where:

$$l_{P1}+l_{P2}+...+l_{P256}=1,l_{Q1}+l_{Q2}+...+l_{Q256}=1$$

The similarity is:

$$\sum_{j=1}^{n}\min(l_{Pi}+l_{Qi})$$

## 4. Shape feature extraction

As the shape and texture is a reflection of the strength of gray, color images should be first converted to gray-level images using the following formula (the number of gray-level is 256):

*gray = 0.30 × R + 0.59 × G + 0.11 × B*

Where, gray means gray value (quantified to 256); R, G, B are red, green and blue components.

Moment Invariants is region-based object shape representation. Suppose R is a binary image, the p+q central moments of R form as:

$$\mu_{p,q}=\sum_{(x,y)\in R}(x-x_c)^p(y-y_c)^q$$

$(x_c, y_c)$ is the center of the object. For scale-independent nature, central moments can be standardized as:

$$\eta_{p,q}=\frac{\mu_{p,q}}{\mu_{0,0}^{2}},\gamma=\frac{p+q+2}{2}$$

Based on these moments, Hu bring forward seven moments of transformations, rotation and scale independence:

$$\Phi_1=\mu_{2,0}+\mu_{0,2}$$

$$\Phi_2=(\mu_{2,0}-\mu_{0,2})^2+4\mu_{1,1}{}^2$$

$$\Phi_3=(\mu_{3,0}-3\mu_{1,2})^2+(\mu_{0,3}-3\mu_{2,1})^2$$

$$\Phi_4=(\mu_{3,0}+\mu_{1,2})^2+(\mu_{0,3}+\mu_{2,1})^2$$

$$\Phi_5=(\mu_{3,0}-3\mu_{1,2})(\mu_{3,0}+\mu_{1,2})[(\mu_{3,0}+\mu_{1,2})^2-3(\mu_{0,3}+\mu_{2,1})^2]$$

$$+(\mu_{0,3}-3\mu_{2,1})(\mu_{0,3}+\mu_{2,1})[(\mu_{0,3}+\mu_{2,1})^2-3(\mu_{3,0}+\mu_{1,2})^2]$$

$$\Phi_6=(\mu_{2,0}-\mu_{0,2})[(\mu_{3,0}+\mu_{1,2})^2-(\mu_{0,3}+\mu_{2,1})^2]+4\mu_{1,1}(\mu_{3,0}+\mu_{1,2})(\mu_{0,3}+\mu_{2,1})$$

$$\Phi_7=3(\mu_{2,1}-\mu_{0,3})(\mu_{3,0}+\mu_{1,2})[(\mu_{3,0}+\mu_{1,2})^2-3(\mu_{0,3}+\mu_{2,1})^2]$$

$$+(\mu_{0,3}-3\mu_{2,1})(\mu_{0,3}+\mu_{2,1})[(\mu_{0,3}+\mu_{2,1})^2-3(\mu_{3,0}+\mu_{1,2})^2]$$

It can be easily verified that the above seven moments are invariable for translation, rotation and scale.

According to the formula, we can calculate the seven invariant moment features values of the target areas, constituting the eigenvectors of the shape. Each element of the eigenvectors is not the same size, when measured in the Euclidean distance as similarity, they will have a great deviation. So in order to eliminate the bias, they should be first normalized.

We use $F = [\Phi_1, \Phi_2, ..., \Phi_7]$ to represent feature vector, then we use $P_1, P_2, ..., P_M$ to stand for images in database. For a image $P_1$, the feature vector is $F_1 = [\Phi_{1,1}, \Phi_{1,2}, ..., \Phi_{1,7}]$. So we can get a M × 7 matrix from M images. We can calculate the mean $E_j$ and standard deviation $\sigma_j$, and then use Gaussian normalization method ($\Phi_{i,j} = |\Phi_{i,j} - E_j| / 3\sigma_j$) to normalize the original sequence. These seven normalized moment invariants is required eigenvectors value, which will be automatically stored in the image feature library.

## 5. Texture feature extraction

In 1973, Haralick proposed Gray Level Co-occurrence Matrix (GLCM) method, which is based on the conditional probability density function. This method has been research for a long history, and is currently recognized as an important texture analysis method.

Co-occurrence matrix is a function of distance and direction. In a given direction and distance, we can calculate the symbiotic grayscale pixel i and j, expressed as the number of co-occurrence matrix element $p(i, j \mid d, \theta)$:

$$p(i, j \mid d, \theta) = \frac{p(i, j \mid d, \theta)}{\sum_i \sum_j p(i, j \mid d, \theta)}$$

Haralick has proposed 14 kinds of the GLCM parameters, among which four parameters are mainly used as following:

1) Moment of inertia (contrast):

$$I = \sum_i \sum_j (i - j)^2 P(i, j)$$

Image contrast can be interpreted as the sharpness of the image: the deeper grooves of the image texture, the greater the contrast is.

2) Energy:

$$E = \sum_i \sum_j [P(i, j)]^2$$

Energy is the measure of gray distribution uniformity of image. The coarser the texture is, the more energy it contains.

3) Entropy:

$$H = \sum_i \sum_j [P(i, j)] \log P(i, j)$$

Entropy is a measure of the amount of information of an image. Entropy relates to the texture information. If there is no texture information, the entropy is zero.

4) Correlation:

$$C = (d,\theta) = \frac{\sum_{i,j}(i-\mu_x)(j-\mu_y)P(i,j)}{\sigma_x \sigma_y}$$

Correlation is used to measure the degree of similarity of the elements in GLCM.

$$\mu_x = \sum_i \sum_j iP(i,j), \mu_y = \sum_i \sum_j jP(i,j), \sigma_x^2 = \sum_i \sum_j (i-\mu_x)^2 P(i,j), \sigma_y^2 = \sum_i \sum_j (j-\mu_y)^2 P(i,j)$$

5) Texture similarity measure

Given image Q, I, the difference between the texture features is defined as:

$$S_t(Q,I) = 1 - \sum_{i=1}^{4} \sum_{j=1}^{4} w_{ij} \mid f_{ij}(Q) - f_{ij}(I) \mid$$

Where, i = 1,...,4 represent the four co-occurrence matrix described above, j = 1,...,4 represent the four co-occurrence matrix parameters, $f_{ij}$ is the normalized co-occurrence matrix parameter, $w_{ij}$ is the weight of each parameter. It can be seen, the more similar the two images, the bigger the St is.

## 6. Conclusion

Let N be the number of images returned for the query, R is images associated with the example in the result, M is images associated with the example in test set S. So the retrieval precision and recall rate can be defined as follow:

$$precision = \frac{R}{N}$$

$$recall = \frac{R}{M}$$

We use 228 images to do the experiment, the result is as table.1.

| Item / Method | Precision | Recall | Response time(ms) |
|---|---|---|---|
| Texture Feature | 60.1% | 72.3% | 1373 |
| Color Feature | 53.7% | 65.2% | 824 |
| Shape Feature | 62.2% | 70.8% | 2544 |
| This paper | 79.6% | 88.3% | 3861 |

Table.1 Comparison on Precision, Recall and Response time

As we can see in table.1, the combination of texture, color and shape texture with GrabCut, strongly improved precision, recall, although it has longer response time.

## References


[1] R M Haralick, K Shangmugam, etc. Texture feature for image classification [J].IEEE Transaction on Systems,1973, SMC-3 (6): 768-780

[2]Anne H，Solberg S，Jain A K.Texture Fusion and Feature Selection Applied to SAR Imagety[J].IEEE Transaction on Geoscience and Remote Sensing，1997，35(2)：475-478

[3]B.S. Manjunathi and W.Y. Ma. Texture Features for Browsing and Retrieval of Image Data .IEEE TRANSACTIONS ON PATTERN ANALYSlS AND MACHINE INTELLIGENCE, VOL. 18, NO. 8, AUGUST 1996.837-842

[4]A. W. Smeulders, M. M. Worring, A. Gupta, and R. Jain, Content-Based Image Retrieval at the End of the Early Years, IEEE Trans. Pattern Anal. Machine Intell., vol.22 no.12, pp1349-1380, 2000.

[5] Deng, Y., Manjunath, B.S. Unsupervised segmentation of color-texture regions in images and video[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2001,Vol:23(8),800~ 810